AFFECTIVE STATES: ANALYSIS AND SONIFICATION OF TWITTER MUSIC TRENDS

Kingsley Ash

Leeds Metropolitan University,
Caedmon Hall 220,
Headingley, Leeds, LS6 3QS, United Kingdom
k.ash@leedsmet.ac.uk

ABSTRACT

This paper describes an approach to the sonification of real-time twitter music trend data realized for the ICAD 2012 Sonification Competition: Listening to the World Listening. The paper will discuss the techniques used to create the sonification and the motivations behind them, including details of the data analysis, mapping strategies, visual display and sonification output.

The system analyses the Twitter Music Trends data feed, which aggregates music listening data from Twitter by artist, as well as the Echo Nest REST API to determine the perceived emotional affect and prevailing descriptions of a selection of the latest trending artists. The resulting data is visualized and sonified in real-time to facilitate analysis and generate an appealing visual and auditory display of the resulting data.

Experience with the system suggests that it is successful in allowing users to determine perceived emotional affect and quality for a number of artists simultaneously, and could allow further investigation into the correlation between these factors. The system also generates appealing visual music that reaches beyond the practice of scientific investigation to reach out to a wider audience.

1. INTRODUCTION

The ICAD 2012 sonification competition is inspired by the idea of music (or sound) about listening to music. It is also inspired by the radical changes over the past decade in how we listen to music and how we share our listening activities with others. Adopting the theme 'Listening to the World Listening', it challenges us to explore what we can learn about listening through the analysis and sonification of social media data about listening.

This response to the competition theme explores our relationship with music listening, both through the new media in which we are now able to share and discuss this music (twitter, blogs, online reviews, etc.) as well as through the words used to describe, categorize and discuss music. In particular, the system examines the words we use to describe the emotional affect the music has on us ("this music makes me feel relaxed/bored/uplifted/angry/etc."), and the words we use to subjectively categorise our listening ("this music is good/bad/fast/slow/unique/predictable/etc."), allowing connections and relationships between these areas to be uncovered and experienced. Furthermore, as this analysis is taking place on live data from Twitter, it allows us to view these relationships in real-time, as the artists in question are listened to and discussed on social media.

The system, created in Max/MSP, contains processes for real-time data collection, data analysis, sonification and visualization and all are discussed in detail in this paper.

2. DATA COLLECTION

The data for this sonification system is obtained from the Twitter Music Trends data feed, which returns a list of the top 50 current trending artists on Twitter. The data is updated every two seconds, and each update is acquired by the system in real-time. From this data the system extracts the artist name, score (ranking popularity of artist on Twitter relative to the most popular artist on the list) and Echo Nest id.

Having obtained the latest Twitter trend data, the system then uses the Echo Nest id tags to access additional information about each artist. In order to gather a range of text documents relating to each artist, the Echo Nest REST API is accessed, using bucket terms to return the 15 most recently available news items, blogs and reviews.

Having completed this process, the system can now analyse the current list of trending artists and a selection of articles in which the artist and their work is discussed. This allows the system to determine the perceived emotions and prevailing descriptions of the artists currently being listened to and discussed across the globe.

3. DATA ANALYSIS

3.1. Valence-arousal space

The most widely used model for the representation of emotions is the 2D valence/arousal space, in which valence is represented on the X axis from highly negative to highly positive, and arousal is represented on the y axis, from calming/soothing to excited/agitated [1]. This model has been widely used to determine the apparent mood of music [2,3] and also to form the basis of music recommendation services [4].

The circumplex model of affect places a number of common emotion words in this 2D valence/arousal space, with negative high arousal emotions in the top left, (e.g. anger, tension, etc.), positive high arousal emotions in the top right (e.g. joy, excitement, etc.), negative low arousal emotions in the bottom left (e.g. boredom, depression, etc.) and positive low arousal emotions in the bottom right (e.g. serenity, calmness, etc.) as shown in figure 1.

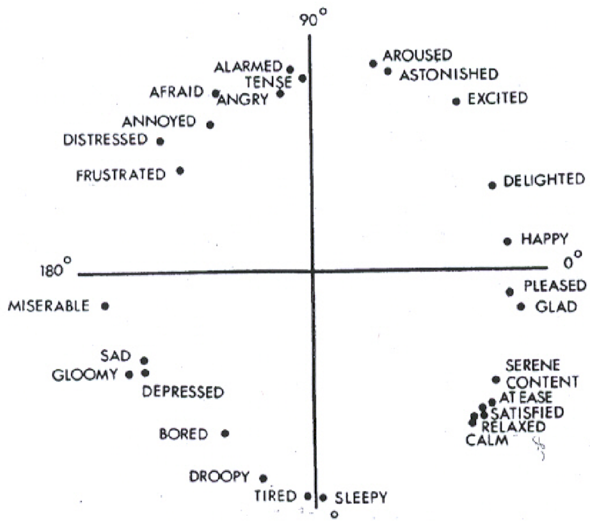

Figure 1: Russell's circumflex model of emotion.

The example emotion words given by Russell and shown in Figure 1 are not always those that come to mind when discussing popular music in the twenty-first century. As a result, a selection of more commonly used emotion words was extracted from a review of online writings referring to popular music artists. Combined with the original emotion words proposed by Russell, a total of 39 words were placed into the valence arousal model (see Figure 2).

High arousal High valence	Low arousal High valence	Low arousal Low valence	High arousal Low valence
Aroused	Pleased	Bored	Confused
Astonished	Glad	Depressed	Frustrated
Stunned	Emotional	Gloomy	Distressed
Excited	Serene	Sad	Annoyed
Inspired	Satisfied	Sorrowful	Painful
Uplifted	Relaxed	Miserable	Aggressive
Surprised	Calm		Angry
Compelled	Soothed		Alarmed
Interested	Dreamy		Tense
Delighted	Sleepy		
Cheerful			
Playful			
Happy			

Figure 2: Emotion words grouped by arousal and valence

For each artist, the system searches through the text obtained from the Echo Nest API to find occurrences of any of these 39 emotion words. Each occurrence is then logged and the total number of each is calculated to give an estimation of the overall current emotional response to each artist.

3.2. Quality-energy space

In order to determine the prevailing descriptions used for each artist, descriptive words are searched for in the same way as above. These descriptive words are also placed in a 2D space, with *quality* running along the X-axis, from bad to good and perceived *energy* along the Y-axis, from slow to fast (see Figure 3).

The number of occurrences of each word is logged, and then the numbers are aggregated for each quadrant of the 2D quality-energy space. These numbers are then combined to give a single measure of quality and a single measure of energy.

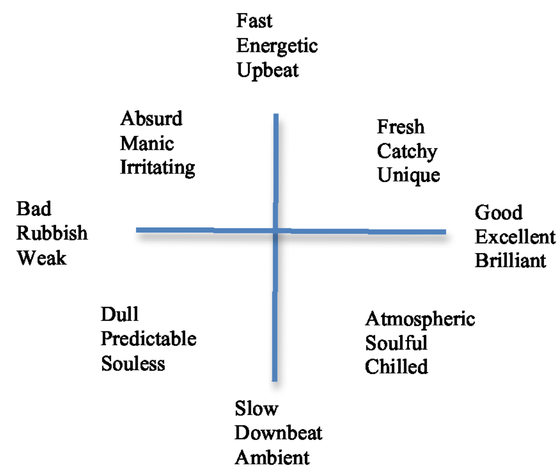

Figure 3: 2D quality-energy space

Having completed the valence-arousal and quality-energy analysis, the results are then used to generate visual and audio outputs for analysis through observation and listening.

4. VISUALISATION

The results of this data analysis are shown in a visualization that simultaneously displays the valence-arousal and quality-energy data for each artist in real-time. A central dot is plotted for each artist in an X-Y space corresponding to their quality-energy score. In this way, the higher the perceived quality of the artist, the further to the right they will appear. Similarly, the higher the perceived energy of an artist, the higher up the visualization space they will appear. As an example, Sigur Ros who are known for their downtempo compositions appear very low down the Y-axis, while the high-energy Japanese pop act AKB48 appear high up the Y-axis.

Having determined the X-Y position of the artist, the emotion words are then plotted around that point at angles corresponding to their position in the 2D valence-arousal model. The length of the lines for each word corresponds to the number of appearances of that word in the text documents. Finally, the name of the artist and their position in the top 50 trending artists is displayed alongside the plot (see Figure 4).

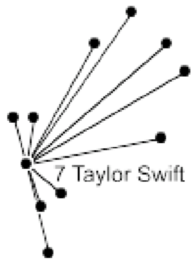

Figure 4: Valence-arousal plot for the artist Taylor Swift.

As can be seen from comparison of Figure 4 and Figure 1, the artist Taylor Swift clearly appears to evoke positive, high arousal emotions such as excitement, though with some elements of positive low arousal states such as calmness and relaxation. The completed plot for each artist therefore shows the arousal-valence scores plotted in an X-Y position that corresponds to the quality-energy scores, allowing all of these parameters to be viewed simultaneously, alongside the artist name and rank.

5. SONIFICATION

Each artist is represented by an individual tone, the timbre of which is synthesized in response to their arousal-valence scores. The fundamental frequency of this sound is determined by the horizontal position on the quality-energy space, and the duration of the sound is determined by the vertical position.

The resulting tone for each artist is synthesised using four components, each corresponding to one quadrant of the valence-arousal model. Positive high arousal emotions (e.g. excitement) are represented using additive synthesis, where the individual scores of each emotion word determine the frequency and amount of the harmonics. Negative high arousal emotions (e.g. annoyance) introduce inharmonic components into the sound, the frequency and amplitude of which are determined by the individual emotion words. Negative low arousal emotions (e.g. tiredness) are represented using filtered noise components in the attack section of the sound, with the centre frequency of each filter determined by the individual scores of each emotion word. Finally, the positive low arousal emotions (e.g. relaxation) are represented by sub-harmonics below the fundamental frequency of the tone. By this method, scores in each quadrant of the valence-arousal space introduce a clearly defined component to the tone that can be picked out in the timbre of the overall sound (see Figure 5).

The overall sound for each artist is played immediately following the collection and analysis of data for that artist. This results in a sonification consisting of a series of consecutive tones allowing both individual and comparative analysis of the data.

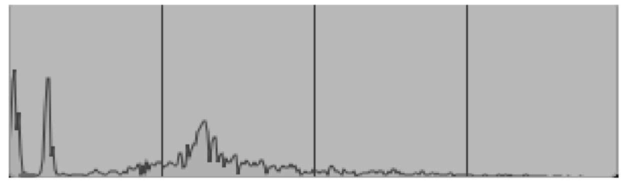

Figure 5: Spectrogram showing fundamental, sub-harmonic and higher frequency noise components of the resulting tone.

6. DISCUSSION

The system successfully enables simultaneous real-time visualization and sonification of music trend data and the results of text-based analysis of recent news, blogs and reviews about each artist. The text analysis process is based on sound foundations in emotional psychology, but there are limitations to the implementation in this system. The analysis looks for individual words, with no regard for the context within which they are used. However, many appearances of a particular word are likely to have some significance, and the combination of a number of words in a similar area of the valence-arousal or quality-energy space are also likely to be significant.

The visualisation is a useful tool to examine the data, allowing both observation of specific points of interest, as well as more general indications of groupings, similarities, trends and movements. However, there is room for further development of this aspect of the system. For example, the visualization doesn't take into account the popularity score of each artist, and this could be utilized as an extra parameter to determine size of the plot for example.

The sonification is also successful in that it allows the listener, with practice, to monitor a large number of simultaneous parameters of the data. The sequential sonification of the data produces repeating patterns in which small variations are easy to detect, but which may turn out to lack interest during repeated listening.

Overall, the system appears to achieve its aim of allowing users to determine perceived emotional affect and quality for a number of artists simultaneously, in a visually and aurally appealing way. Further experience with the system is required to fine-tune the sonification and visualisation algorithms, as well as to fully understand its possibilities and limitations.

7. REFERENCES

[1] J. A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology,* vol. 39, no. 6, pp. 1161-1178, 1980.

[2] D. Lu, L. Liu, and H. Zhang, "Automatic mood detection and tracking of music audio signals," in *IEEE Transactions on Audio, Speech and Language Processing,* vol. 14, no. 1, pp. 5-18, January 2006.

[3] C. Laurier, M. Sordo, J. Serra, and P Herrera, "Music mood representations from social tags," in *Proc. of Int. Soc. For Music Information Retrieval Conf. (ISMIR)*, Kobe, Japan, 2009.

[4] J. Kim, S. Lee, S. Kim, and W. Yoo, "Music mood classification model based on arousal-valence values," in *Proc. Of Conf. of Advanced Communication Technology (ICACT),* Seoul, Korea, 2011, pp. 292-295.